Volume 14 Number 12 1986 Nucleic Acids Research

Analysis of the nucleotide sequence of the *ereB* gene encoding the erythromycin esterase type II

Michel Arthur, Denise Autissier+ and Patrice Courvalin*

Unité des Agents Antibactériens, CNRS U.A. 271, Institut Pasteur, 28, rue du Dr Roux, 75724 Paris Cedex 15 and +Centre de Recherche Roussel-Uclaf, 93230 Romainville, France

Received 23 April 1986; Accepted 23 May 1986

ABSTRACT

We have determined the nucleotide sequence of the ereB gene of plasmid pIP1527 which confers high-level resistance to erythromycin by inactivation in Escherichia coli. The open reading frame of the ereB gene, 1257-bp, was defined by initiation and termination codons and by cloning in vitro. The corresponding protein has a calculated M_r of 48,118 in close agreement with a previous estimation, 51,000, by electrophoresis of minicell extracts in SDS-polyacrylamide gels. The structure of the modified erythromycin was determined by physico-chemical techniques including mass spectrometry, infrared spectrophotometry and ^{13}C nuclear magnetic resonance. The data obtained indicated that like ereA (Ounissi and Courvalin, 1985) ereB encodes an erythromycin esterase . Comparison of the amino acid sequences of the two isozymes did not reveal any statistically significant homology. Analysis of the nucleotide sequence of the ereB gene suggests that this resistance determinant should be exogenous to E. coli.

INTRODUCTION

Macrolide-lincosamide-streptogramin (MLS) antibiotics are mainly active against Gram-positive bacteria. In general, acquired resistance towards MLS antibiotics involves N^6-dimethylation of a specific adenine residue in 23S ribosomal RNA. The modified ribosome binds the drugs less efficiently which leads to a co-resistance phenotype towards these chemically unrelated compounds (1).

Enterobacteria, like most Gram-negative organisms, are intrinsically resistant to low levels of MLS antibiotics which precludes their therapeutical use in systemic infections (2). However, because of the high local concentrations achieved (2), erythromycin has recently found therapeutical applications in the modulation of the Gram-negative flora of the intestinal tract (2,3). The intestinal carriage of enterobacteria highly resistant to erythromycin (MIC $\geq$ 500 µg/ml) is frequent but is usually associated with previous intake of the drug (3).

We have recently described a new resistance phenotype in a clinical isolate of E. coli to high levels of erythromycin alone (4). This resistance is due to the synthesis of an erythromycin esterase which hydrolyzes the lactone ring of the 14 - membered

macrolides erythromycin and oleandomycin (5). The plasmid-borne gene encoding the enzyme, designated ereA, was sequenced (6) and its distribution in clinical isolates of enterobacteria highly resistant to erythromycin was studied by colony hybridization using an intragenic probe (7). Our results indicated the spread of the ereA gene but also a polymorphism, at the DNA level, for the erythromycin modifying enzyme.

One of the strains, BM2570, which did not hybridize with the ereA probe was found to be resistant to erythromycin by two distinct mechanisms and the corresponding plasmid genes were cloned separately by recombinant DNA techniques (8). The first gene, ermA, confers resistance to MLS antibiotics in the absence of inactivation . Nucleotide sequence comparisons indicated that this phenotype is due to recent acquisition by E. coli of a gene closely related to the ermAM gene encoding the RNA methylase in streptococci (9). The second gene, named ereB, codes for an enzyme with an apparent M_r of 51,000 which confers resistance to erythromycin by inactivation (8). We have sequenced ereB and determined the biochemical mechanism by which it confers resistance to erythromycin.

MATERIALS AND METHODS

Bacterial strains, bacteriophages and plasmids.

Fragments of DNA to be sequenced were transfected into E. coli JM101 (10) using M13mp8 or M13mp9 (11) bacteriophage vectors. Recombinant plasmids were introduced in E. coli strain BM694 (12) by transformation. The sources and properties of the plasmids used in this study are listed in Table 1.

Media.

Brain Heart Infusion broth and agar (Diagnostics Pasteur) were used. Disc sensitivity tests were done on Mueller-Hinton agar. All incubations were at 37°C.

Inactivation of erythromycin by resting cells.

Inactivation of erythromycin by resting cells of E. coli BM694/pAT63 or BM694/pAT72 in 0.1 M phosphate buffer (pH7.0) was as described (5).

Extraction and purification of inactivated erythromycin.

The products of inactivation of erythromycin were recovered according to the purification procedure of Barthélémy et al. (5). The modified compounds were isolated as a methylene chloride soluble fraction designated fraction B.

Identification of the modified compounds.

Mass spectrometry : Mass spectrometry measurements were obtained by field desorption on a Varian MAT 311A apparatus.

NMR : Natural abundance of ^{13}C NMR spectra were obtained on a Brucker 250 WM apparatus.

IR spectra : IR absorption spectra were obtained on a Perkin Elmer 580 apparatus.

Preparation of DNA.

Large-scale isolation of pBR329, pUC8, and derivative plasmids DNA was as described (13).

Electrophoresis and purification of DNA restriction fragments.

The DNA restriction fragments were separated by electrophoresis in horizontal slab gels (20 x 15 x 0,7 cm) containing 0.8 % low-temperature-gelling agarose Type VII (Sigma) . DNA fragments were purified as described (14).

DNA sequencing.

Restriction DNA fragments were cloned in bacteriophages M13mp8 and M13mp9 and sequenced by the chain terminator technique (15). The complete DNA sequence was arranged using DBCOMP and DBUTIL computer programs (16).

DNA stability profile.

The thermodynamic properties of the DNA in ereB were predicted from the DNA sequence of the insert in pAT72 according to the algorithm of Gabarro-Arpa and Michel (17). The stability profile is defined as the plot of the DNA stability parameter P versus the DNA sequence at a constant environment parameter W (18). The parameter P has no dimension and ranges between 0 (minimum stability) and 1 (maximum stability). An excellent fit of the experimental melting profiles with those calculated from the sequence is observed within the W range that is experimentally accessible (W from 5 to 10 corresponding to ionic strength from 10^{-1} to 10^{-3} M monovalent ions). However, it is admitted that this correlation remains valid up to W = 2.5 (18).

Comparison of amino acid sequences.

The amino acid sequences were compared using a computer and the algorithm of Wilbur and Lipman (19). The statistical significance of the homologies between the amino acid sequences were tested with the algorithm of Sellers (20) as generalized (21,22).

Enzymes and reagents.

Restriction endonucleases, DNA polymerase I (large fragment), T4 DNA ligase, calf alkaline phosphatase (Boehringer Mannheim) were used according to the manufacturer's recommendations. Lysozyme was provided by Sigma, Inc. Deoxyadenosine 5'-[α-^{32}P] triphosphate, triethylammonium salt, was from the Radiochemical Center, Amersham. Deoxynucleoside triphosphates, dideoxynucleoside triphosphates and bacteriophages M13mp8 and M13mp9 RF DNA were from PL-Biochemicals. M13 pentadecamer primer was purchased from Biolabs. Erythromycin base was provided by Roussel-Uclaf.

RESULTS

Structure of the modified erythromycin.

The substrate profiles and the RF values of the detoxification products of the

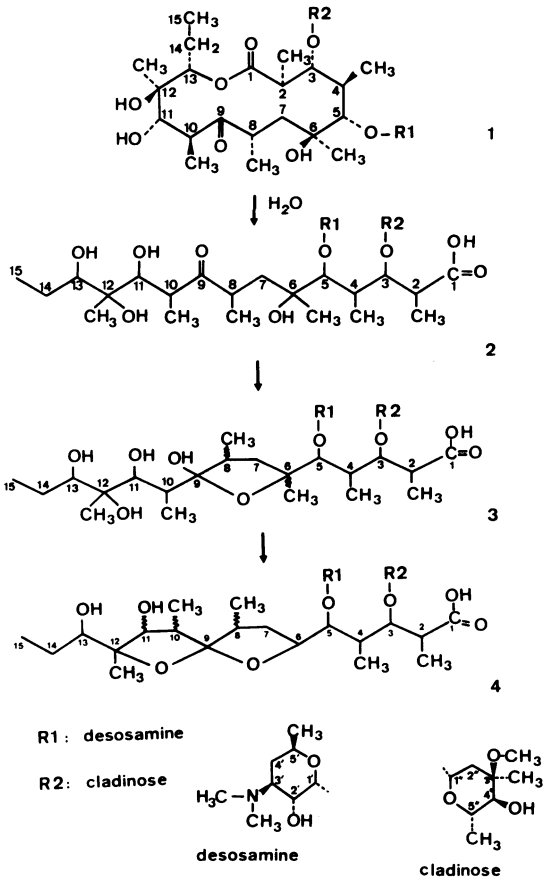

Figure 1. Reaction catalyzed by the erythromycin esterases types I and II. Enzymic hydrolysis of the lactone ring of erythromycin is followed by the formation of an hemiaketal by internal condensation and dehydratation. Compound 4 is the major end product of detoxification of erythromycin by resting cells at pH7.0 (5).

reactions catalyzed by the erythromycin esterase (5) and the enzyme encoded by the ereB gene are indistinguishable (8). We therefore established in parallel the structure of the two modified compounds using the strategy described by Barthélémy et al. (5). Erythromycin was inactivated by resting cells and the modified antibiotic was isolated as a methylene chloride soluble fraction designated fraction B. After dissolution in chloroform the IR spectra of the two fractions B were indistinguishable. We noted the absence of ketone and lactone in both spectra. There was no difference between the ^{13}C NMR spectra of the two fractions B. Mass spectra showed a molecular peak at 734 (erythromycin 734) for both compounds. The interpretations of NMR and mass spectra

Table 1. **Plasmids and their origins**

Plasmid	Relevant characters	Origin	Reference
pAT63	Tra⁻ Mob⁻ ereA Ap(a)	pBR322Ω[pIP1100 partial Sau3A(ereA)-1.66kb]	Ounissi and Courvalin (6)
pAT70	Tra⁻ Mob⁻ ereB Cm Tc	pBR329Ω[pIP1527BamHI (ereB)-3.8kb]	Arthur and Courvalin (8)
pAT72	Tra⁻ Mob⁻ ereB Ap	pUC8Ω[pAT70 partial Sau3A (ereB)-1906bp]	Arthur and Courvalin (8)
pUC8	Tra⁻ Mob⁻ Ap	In vitro construction	Vieira and Messing (38)

(a) Nomenclature of phenotypic characters of plasmids is according to Novick et al. (39)

of fraction B were already provided (5). Our results are in agreement with the proposed formation of an hemiketal by internal condensation and dehydratation after enzymic hydrolysis of the lactone ring of erythromycin yelding to the formation of compound 4 (Fig. 1). We concluded that like ereA, ereB encodes an erythromycin esterase. The two

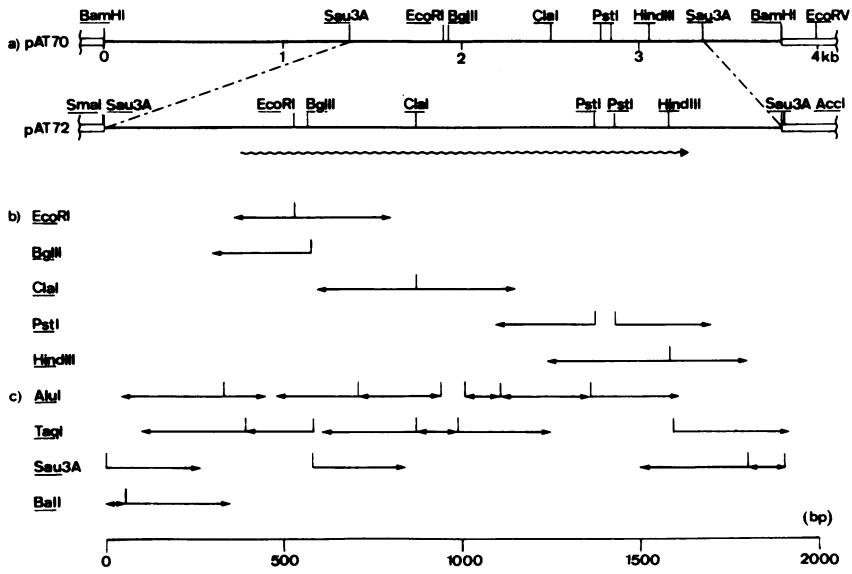

Figure 2. a) Restriction maps of the inserts in plasmids pAT70 and pAT72. Vector and insert segments are depicted by open and shaded segments, respectively. Wavy arrow represents the ereB ORF. Sizes are specified in kb. b) and c) Nucleotide sequence strategy of part of the insert in pAT70 and of the insert in pAT72, respectively. Arrows indicate the direction and extent of the dideoxy sequencing reaction. Size are specified in bp.

isozymes will be referred to as erythromycin esterases type I and II, respectively.

Nucleotide sequence of the insert in pAT72.

Plasmids pAT70 and pAT72 (Table 1) were already described (8). Briefly, plasmid pAT70 consists of a 3.8-kb BamHI DNA fragment of the natural plasmid pIP1527 conferring resistance to erythromycin cloned into pBR329. The ereB gene was subcloned at the BamHI restriction site of plasmid pUC8 as part of a 1.9-kb Sau3A partial digest of the pAT70 insert DNA. The resulting hybrid, pAT72, does not possess a BamHI restriction site (Fig. 2a).

The purified 3.8-kb BamHI fragment of plasmid pAT70 was digested independently with BglII, ClaI, EcoRI, HindIII, or PstI endonuclease and the resulting DNA fragments were cloned in the replicative forms of M13mp8 or M13mp9 and in pUC8. None of these restriction fragments was found to express erythromycin resistance after cloning in pUC8 (data not shown). Nucleotide sequence data were obtained by the chain terminator technique (Fig. 2b). In order to complete the nucleotide sequence, the 1.9-kb AccI-SmaI fragment of plasmid pAT72 was purified, digested with the endonucleases listed in Fig. 2c and cloned in the replicative forms of M13mp8 or M13mp9. In each experiment, specific clones were identified by the dideoxy-T screening method and sequenced. The entire nucleotide sequence of the 1906-bp fragment of pAT72 obtained by computer analysis is shown in Fig. 3. Since the insert in pAT72 is derived from that in pAT70 following partial Sau3A digestion we sequenced the pAT70 DNA fragments which span the two Sau3A sites in the ereB sequence to ensure that ligation of Sau3A fragments in an aberrant order did not occur during pAT70 construction.

Identification of the gene coding for erythromycin esterase.

The localization of the initiation (ATG, GTG) and the termination (TAG, TGA, TAA) codons in the three reading frames on each strand of DNA indicated a single open reading frame (ORF) for the esterase. This ORF of 1257-bp extends from the ATG codon at position 383 to the termination codon TAA at position 1640. This TAA codon is followed by a second termination codon, TGA, at position 1643. Six bp upstream of this ATG codon is located a putative ribosome binding site (RBS) sequence (Fig. 3). This sequence, ACAGGAGG, is complementary for 6 out of 8 bases (underlined) with the 3'OH terminus of the 16S rRNA (3'-OH AUUCCUCC5') of E. coli. The free energy of interaction (ΔG) of the most stable structure between the putative RBS sequence and the 3'-OH terminus of the 16S rRNA, calculated as described (23), is -16.6 kcal mol^{-1}. The nucleotide sequence which extends from the ATG codon at position 383 to the TAA codon at position 1640 can code for a protein of 419 amino acids with an M_r calculated on the basis of the amino acid composition of 48,118. The codon usage in this ORF is presented in Table 2.

Upstream from the coding sequence for the erythromycin esterase, there is a small ORF extending from the ATG codon at position 177 to the two contiguous

```
GATCGTAAAG CGTATTCTAC GGCAATACGG CTACCCACCA GATATGCAAA TGCTGGCCAC                60
GGAAACCGTT CTCAAGCAAG CAGAATTAAT TGCTGAGGAA TTGACCCATG AGAATTAATC               120
                      -35                       -10
GGAAAACCGC CTACGCACGT CCATGTGCTC TGGCAGAAGA GGGGCGAGGT TAAGAAATGA               180
                                          RBS
AGACAAAAAC CAAAGAGGCA CTCTGCCCAT CGCACAACAG CGGCTATATG GCTACGCTTC               240
CGTCCGCCCA AATTGGCTTC CTTCGGTCGC CAACTTCGGG TACTCGCGGA ACGTTGGGTG               300
CAATTAGCCC GCGGGGAGCA TTGGCTTTAA GAGCTGCGCA TCGTGCAGCA AAGAAATTGG               360

                                  MET ARG PHE GLU GLU TRP VAL LYS ASP LYS HIS ILE PRO
    AAACGAATAC AGGAGGAAAAATA ATG AGG TTC GAA GAA TGG GTC AAA GAT AAG CAT ATT CCT   421
              RBS

PHE LYS LEU ASN HIS PRO ASP ASP ASN TYR ASP ASP PHE LYS PRO LEU ARG LYS ILE ILE
TTC AAA CTG AAT CAC CCT GAT GAT AAT TAC GAT GAT TTT AAG CCA TTA AGA AAA ATA ATT   481

GLY ASP THR ARG VAL VAL ALA LEU GLY GLU ASN SER HIS PHE ILE LYS GLU PHE PHE LEU
GGA GAT ACC CGA GTT GTA GCA TTA GGT GAA AAT TCT CAT TTC ATA AAA GAA TTC TTT TTG   541

LEU ARG HIS THR LEU LEU ARG PHE PHE ILE GLU ASP LEU GLY PHE THR THR PHE ALA PHE
TTA CGA CAT ACG CTT TTG CGT TTT TTT ATC GAA GAT CTA GGT TTT ACT ACG TTT GCT TTT   601

GLU PHE GLY PHE ALA GLU GLY GLN ILE ILE ASN ASN TRP ILE HIS GLY GLN GLY THR ASP
GAA TTT GGT TTT GCT GAG GGT CAA ATC ATC AAT AAC TGG ATA CAT GGA CAA GGA ACT GAC   661

ASP GLU ILE GLY ARG PHE LEU LYS HIS PHE TYR TYR PRO GLU GLU LEU LYS THR THR PHE
GAT GAA ATA GGC AGA TTC TTA AAA CAC TTC TAT TAT CCA GAA GAG CTC AAA ACC ACA TTT   721

LEU TRP LEU ARG GLU TYR ASN LYS ALA ALA LYS GLU LYS ILE THR PHE LEU GLY ILE ASP
CTA TGG CTA AGG GAG TAC AAT AAA GCA GCA AAA GAA AAA ATC ACA TTT CTT GGC ATT GAT   781

ILE PRO ARG ASN GLY GLY SER TYR LEU PRO ASN MET GLU ILE VAL HIS ASP PHE PHE ARG
ATA CCC AGA AAT GGA GGT TCA TAC TTA CCA AAT ATG GAG ATA GTG CAT GAC TTT TTT AGA   841

THR ALA ASP LYS GLU ALA LEU HIS ILE ILE ASP ASP ALA PHE ASN ILE ALA LYS LYS ILE
ACA GCG GAT AAA GAA GCA CTA CAC ATT ATC GAT GAT GCA TTT AAT ATT GCA AAA AAG ATT   901

ASP TYR PHE SER THR SER GLN ALA ALA LEU ASN LEU HIS GLU LEU THR ASP SER GLU LYS
GAT TAC TTC TCC ACA TCA CAG GCA GCC TTA AAT TTA CAT GAG CTA ACA GAT TCT GAG AAA   961

CYS ARG LEU THR SER GLN LEU ALA ARG VAL LYS VAL ARG LEU GLU ALA MET ALA PRO ILE
TGC CGT TTA ACT AGC CAA TTA GCT CGA GTA AAA GTT CGC CTT GAA GCT ATG GCT CCA ATT  1021

HIS ILE GLU LYS TYR GLY ILE ASP LYS TYR GLU THR ILE LEU HIS TYR ALA ASN GLY MET
CAC ATT GAA AAA TAT GGG ATT GAT AAA TAT GAG ACA ATT CTG CAT TAT GCC AAC GGT ATG  1081

ILE TYR LEU ASP TYR ASN ILE GLN ALA MET SER GLY PHE ILE SER GLY GLY GLY MET GLN
ATA TAC TTG GAC TAT AAC ATT CAA GCT ATG TCG GGC TTT ATT TCA GGA GGC GGA ATG CAG  1141

GLY ASP MET GLY ALA LYS ASP LYS TYR MET ALA ASP SER VAL LEU TRP HIS LEU LYS ASN
GGC GAT ATG GGT GCA AAA GAC AAA TAC ATG GCA GAT TCT GTG CTG TGG CAT TTA AAA AAC  1201

PRO GLN SER GLU GLN LYS VAL ILE VAL VAL ALA HIS ASN ALA HIS ILE GLN LYS THR PRO
CCA CAA AGT GAG CAG AAA GTG ATA GTA GTA GCA CAT AAT GCA CAT ATT CAA AAA ACA CCC  1261

ILE LEU TYR ASP GLY PHE LEU SER CYS LEU PRO MET GLY GLN ARG LEU LYS ASN ALA ILE
ATT CTG TAT GAT GGA TTT CTA AGT TGC CTA CCA ATG GGC CAA AGA CTT AAA AAT GCC ATT  1321

GLY ASP ASP TYR MET SER LEU GLY ILE THR SER TYR SER GLY HIS THR ALA ALA LEU TYR
GGT GAT GAT TAT ATG TCT TTA GGT ATT ACT TCT TAT AGT GGG CAT ACT GCA GCC CTC TAT  1381

PRO GLU VAL ASP THR LYS TYR GLY PHE ARG VAL ASP ASN PHE GLN LEU GLN GLU PRO ASN
CCG GAA GTT GAT ACA AAA TAT GGT TTT CGA GTT GAT AAC TTC CAA CTG CAG GAA CCA AAT  1441

GLU GLY SER VAL GLU LYS ALA ILE SER GLY CYS GLY VAL THR ASN SER PHE VAL PHE PHE
GAA GGT TCT GTC GAG AAA GCT ATT TCT GGT TGT GGA GTT ACT AAT TCT TTT GTC TTT TTT  1501

ARG ASN ILE PRO GLU ASP LEU GLN SER ILE PRO ASN MET ILE ARG PHE ASP SER ILE TYR
AGA AAT ATT CCT GAA GAT TTA CAA TCC ATC CCG AAC ATG ATT CGA TTT GAT TCT ATT TAC  1561

MET LYS ALA GLU LEU GLU LYS ALA PHE ASP GLY ILE PHE GLN ILE GLU LYS SER SER VAL
ATG AAA GCA GAA CTC GAG AAA GCT TTC GAT GGA ATA TTT CAA ATT GAA AAG TCA TCT GTA  1621

SER GLU VAL VAL TYR GLU *** ***
TCT GAG GTC GTT TAT GAA TAA TGA AATAG CCAGTATAAC TTAACTAAAG ACCGCCATCC          1680

ATGGCGAGAG AAGCCAATGC GAGAGCGGGC TAACAGACGC GCGTATAGGA AGTTCAAACG                1740

CCCTTTTCGG GCAGTCTGCT GGGTAGGCGG CGGTCGCGCA AGCCCCGTTT TGGGCACGGA                1800

TCGGACGTCT GTGAGTGGGA TTTCGGCATC GCGGGGCCAC GCAGCGGCGT TGTCAGCAGC                1860

CATGCTTCGC TGATTCCCGA CAGCGGGCCG AGCGCGCCTG CGGATC                               1906
```

Figure 3. Nucleotide sequence of the 1906-bp Sau3A insert of plasmid pAT72 and deduced amino acid sequence of the erythromycin esterase type II. Numbering begins at the Sau3A site. Presumed -35 recognition site, -10 Pribnow box, and RBSs are underlined. Homologies in the two hexamers with the consensus sequence of E. coli promoters are indicated by bold lettering. Bases complementarity to the 3'-OH terminus of E. coli 16S rRNA are indicated by bold lettering. Arrows indicate the 9 bp direct repeats.

Table 2. **Codon usage in the ereB structural gene**

Phe	TTT	22	Ser	TCT	11	Tyr	TAT	12	Cys	TGT	1
	TTC	9		TCC	2		TAC	7		TGC	2
				TCA	4						
				TCG	1	***	TAA	1	***	TGA	1
Leu	TTA	12					TAG	0			
	TTG	3									
			Pro	CCT	3	His	CAT	11	TRP	TGG	4
				CCC	2		CAC	4			
				CCA	7				Arg	CGT	2
Leu	CTT	4		CCG	2					CGC	1
	CTC	3				Gln	CAA	10		CGA	5
	CTA	7					CAG	4		CGG	0
	CTG	5	Thr	ACT	6						
				ACC	2	Asn	AAT	14	Ser	AGT	3
Ile	ATT	21		ACA	8		AAC	6		AGC	1
	ATC	6		ACG	2						
	ATA	9				Lys	AAA	25	Arg	AGA	6
			Ala	GCT	8		AAG	4		AGG	2
Met	ATG	12		GCC	4						
				GCA	13	Asp	GAT	25	Gly	GGT	12
Val	GTT	6		GCG	1		GAC	4		GCC	6
	GTC	4								GGA	9
	GTA	5				Glu	GAA	19		GGG	2
	GTG	3					GAG	11			

termination codons TAATGA at position 381 and 384. The initiation codon for the erythromycin esterase at position 383 (underlined) spans these two termination codons. In this 207-bp ORF, the ATG codon at position 177 is preceeded by an RBS-like sequence GGGGCGAGG, which possesses poor complementarity (underlined) with the 3'-OH terminus of E. coli 16S rRNA (ΔG= -9.4 kcal mol^{-1}).

Comparison of the amino acid sequences of the erythromycin esterases.

The amino acid sequences of the erythromycin esterases of type I and II were compared. An alignment obtained with the algorithm of Wilbur and Lipmann (19) is shown in Fig. 4. In this alignment, identical amino acids were found in 71 positions (17%) and, by adding homologies per classes, in 126 positions (31%). These percentages are calculated with respect to the length of the type I enzyme, including the first termination codon. However, a cluster of homology, from position 279 to 309 (erythromycin esterase type II numbering), contains seventeen out of thirty-one identical amino acids . To test the statistical significance of these homologies we constructed alignments between shuffled sequences by the Sellers's algorithm (20-22)(data not shown). We concluded that the amino acid sequences of the two erythromycin esterases do not share statistically significant homology.

Base composition and stability profile of the insert in pAT72.

The guanosine plus cytosine mol % (GC content) of the insert in pAT72 is 42%, a

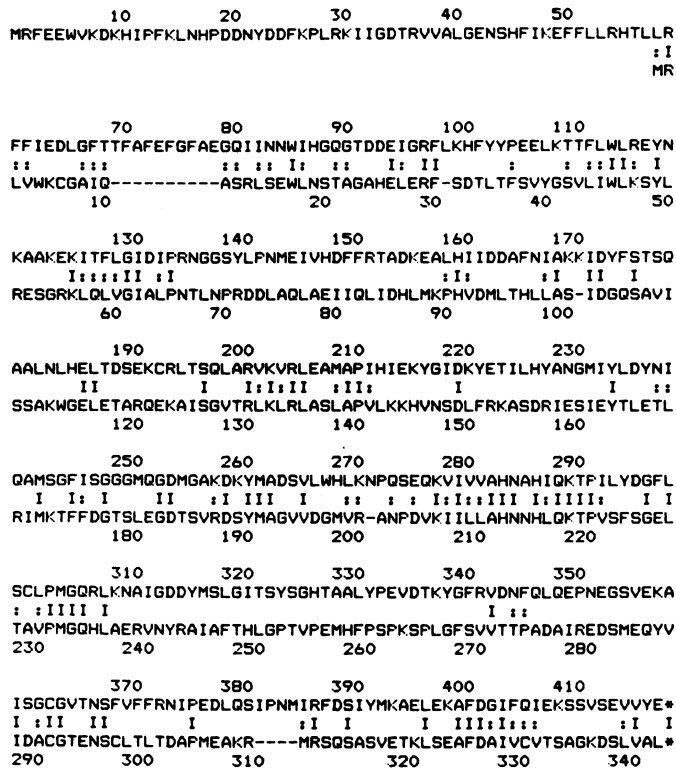

Figure 4. Comparison of the amino acid sequences for the two types of erythromycin esterase. Alignment was obtained by the Wilbur and Lipman algorithm (17). We took into account homologies between amino acids belonging to the same class. Grouping of amino acids in classes was as follows : [RK], [LFPMVI], [STQNC], [AGW], [H], [ED], [Y]. The dot matrix was filtered to only show matches of length ≥ 2, we used a window size of 20 and a gap penalty of 6. Strict homologies are indicated by the letter I and matches between two amino acids belonging to the same class by two dots. Lower and upper numbering are that of the erythromycin esterases type I and II respectively.

value significantly lower than that (50%) of the E. coli genome (24) and varies along the sequence. Upstream from the initiation codon for the erythromycin esterase (from positions 1 to 382) the GC content is 51%. The ORF for the erythromycin esterase (positions 383-1639) has a poor GC content (36%) whereas the 3'OH terminal part of the sequence (1640-1906) is rich in GC (61%).

Because of the non random distribution of the GC content in ereB we determined the stability profile of the insert in pAT72 (Fig. 5). As expected from its overall GC content, the ORF is situated in the most unstable part of the sequence. Interestingly, this unstable region (from positions 350 to 1669) is almost perfectly limited to the translated part of the sequence, starting 19 bp upstream from the RBS sequence and

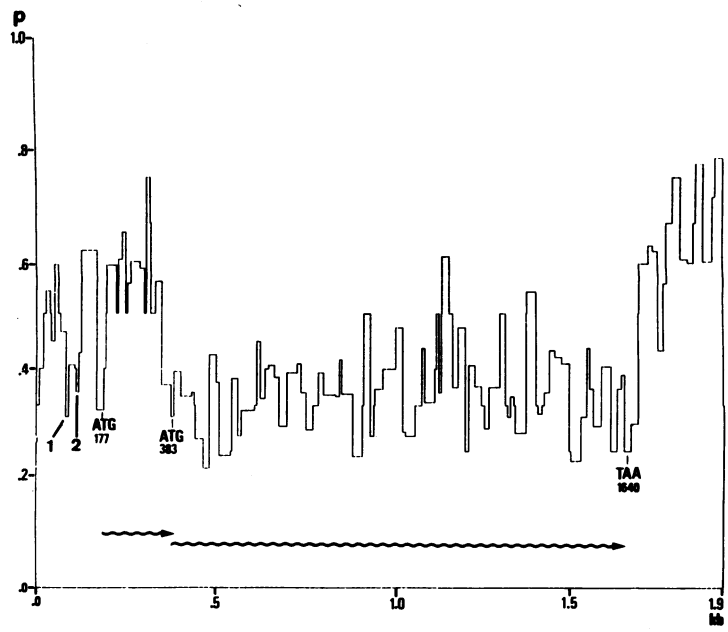

Figure 5. Stability profile of the DNA of the insert in pAT72. The stability profile is the plot of the DNA stability parameter P versus the DNA sequence. The value of environmental parameter was W = 2.5. The short and long wavy arrows indicate the small and ereB ORFs, respectively. The corresponding ATG at positions 177 and 383 are indicated. The two (1 and 2) unstable domains are located in the region corresponding to the proposed promoter. Domain 1 (position 84 to 91) contains the first of the two 9-bp direct repeats and part of the presumed -35 recognition site. Domain 2 (position 113 to 119) contains the 3'-OH terminus of the second 9-bp direct repeats and the presumed -10 Pribnow box.

ending 29 bp after the termination codon. We examined the stability profile for thermodynamic signals tentatively associated with the initiation of translation and transcription (18). The two proposed initiator codons at positions 177 and 383 were situated in domains of low stability (P<0.4), a feature characteristic of the initiator codons in procaryotes (25, 26). A Computer search for literal homology with the -35 TTGACA and -10 TATAAT consensus sequences known to be implicated in the initiation of transcription in E. coli (27) did not reveal any well conserved hexamer. However, comparison of the stability profile of the DNA sequence upstream from the ATG codon for the erythromycin esterase with those of well characterized E. coli promoters (28) indicated a single promoter-like sequence (Fig. 5). From positions 84 to 91 and 113 to 119 two unstable domains were separated and were surrounded by stable domains. A search in these domains for literal homology with the consensus sequence of E. coli promoters revealed two poorly conserved hexamers (Fig. 3). These sequences overlap with a perfect 9 bp direct repeat (Fig. 3).

DISCUSSION

We have determined the nucleotide sequence of ereB from plasmid pIP1527 which confers high level resistance to erythromycin by inactivation. Determination, by physico-chemical techniques, of the structure of the modified antibiotic indicated that ereB, like ereA (5,6), encodes an erythromycin esterase which hydrolyzes the lactone ring of the antibiotic. Analysis of the nucleotide sequence of the 1906-bp insert of plasmid pAT72 which expresses erythromycin resistance (Fig. 3) revealed a single ORF for the esterase. This ORF is preceded by a putative RBS sequence which is complementary to 6 out of 8 bases of the 3'OH end of the nucleotide sequence of E. coli 16S ribosomal RNA. The corresponding protein has a calculated M_r of 48,118 in agreement with the previous estimation of 51,000 by electrophoresis of minicell extracts in polyacrylamide gels (8). We believe that the ATG codon at position 383 is most likely the initiation codon for the esterase, since other in frame ATG or GTG codons in the ORF are not preceded by a RBS-like sequence and since they would also initiate proteins of a calculated M_r lower than that estimated by up to 39%. Moreover, the restriction sites recognized by endonucleases BglII, ClaI, EcoRI, HindIII and PstI situated within the proposed ORF were found to be located in sequences necessary for the expression of erythromycin resistance (Fig. 2). Search for stability profiles (28) and literal homologies (27) with well characterized E. coli promoters showed a putative promoter-like sequence (Fig. 3 and 5). We do not know the biological role, if any, of the 9-bp direct repeat (Fig. 3) which overlapped the sequence of this putative promoter. Erythromycin resistance encoded by ereB is expressed constitutively (8).

We compared the erythromycin esterases type I and II by constructing alignments between the two amino acid sequences. The low percentage of homology (17%) found (Fig. 4) indicated the absence of a close structural relationship between the two proteins. However, the nature of the amino acids situated in a cluster of homology between positions 279 to 309 (erythromycin esterase type II numbering, Fig. 4) suggests that they may play a role in the substrate binding or catalytic site of the two isozymes. The lack of statistically significant homology between the amino acid sequences of the enzymes and their difference in size, 344 and 419 residues, might be indicative of a convergent evolution of unrelated genes rather than a divergent evolution from a common ancestor.

The GC content of ereB (36%) is significantly different from that of E. coli genome (50%) (24). The codon usage in E. coli (29, 30) and in ereB (Table II) appears to be different. Preferential codon usage, which correlates to the GC content, is assumed to be a genome strategy contrary to amino acid usage in proteins (31-34). In order to evaluate the respective roles of amino acid composition and codon usage in the GC content of ereB we distinguished two types of bases/position in this ORF. In the first type of position, mutation (A or T$\rightleftharpoons$G or C) would result in a change of the amino acid

encoded by the corresponding codon. In the second type of position, such mutations would be silent. The GC content in the first type of bases was 40 % whereas this proportion was lower, 29 %, in the second type of position. This result indicates a strong preference for AT rich codons among codons specifying the same amino acid. The difference between the GC content of ereB and that of E. coli genome is also partly due, but at a lesser extent, to a high proportion of amino acids specified by AT rich codons in the erythromycin esterase type II. The low GC content of ereB is therefore mostly due to a specific codon usage (Table II) which is different from that of E. coli. This observation strongly suggests an exogenous origin of ereB.

The ermA gene specifies MLS resistance and is physically linked to ereB on plasmid pIP1527 (8). Genes ermA and ereB are transcribed convergently from opposite promoters (this paper and ref. 8,9). The termination codons of the two corresponding proteins are approximately 3.6 kb apart. The GC content of ermA is also different from that of E. coli. The confirmation that ermA was exogenous to E. coli came from the finding that part of the nucleotide sequence of this gene of plasmid pAM77 (35) and of transposons Tn1545(36) and Tn917 (37) from streptococci were identical (9). The GC content of ermA is similar to that of streptococci suggesting a transfer under natural conditions of a gene from Gram-positive cocci to E. coli (8). Despite their physical linkage in pIP1527 and a similar GC content, a common origin in Gram-positive cocci for ereB and ermA remains to be demonstrated since inactivation of erythromycin was never reported in these bacteria.

The analysis of the nucleotide sequence of ermA suggested that the GC content of nucleotide sequences can be used to identify DNA from exogenous origin. Interestingly, the ORF for the erythromycin esterase type II is surrounded by GC rich sequences (Fig. 5) and we found a small ORF preceded by a RBS-like sequence upstream from ereB (Fig. 3). It is therefore tempting to assume that exogenous DNA is limited to the ORF for the esterase and has been integrated into a gene of pIP1527.

ACKNOWLEDGEMENTS

We thank C. Reiss for computer analysis of the stability profile of DNA and for helpful discussions. M. Arthur was recipient of a Roussel-Uclaf fellowship.

*To whom correspondence and reprint requests should be addressed

REFERENCES

1. Weisblum, B. (1985) J. Antimicrob. Chemother. 16, Suppl.A, 63-90.
2. Andremont, A. and Tancrède, C. (1981) Ann. Microbiol. (Inst. Pasteur) 132B, 419-427.
3. Andremont, A., Sancho-Garnier, H. and Tancrède, C. (1986) Antimicrob. Agents Chemother. (submitted).

4. Andremont, A., Gerbaud, G. and Courvalin, P. (1986) Antimicrob. Agents Chemother. 29, 515-518.
5. Barthélémy, P., Autissier, D., Gerbaud, G. and Courvalin, P. (1984) J. Antibiotics 37, 1692-1696.
6. Ounissi, H. and Courvalin, P. (1985) Gene 35, 271-278.
7. Arthur, M., Andremont, A. and Courvalin, P. (1986) Ann. Microbiol. (Inst. Pasteur) 137A (in press).
8. Arthur, M. and Courvalin, P. (submitted).
9. Arthur, M. and Courvalin, P. (submitted).
10. Messing, J. (1979) Tech. Bull. 1, 43-44.
11. Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
12. Labigne-Roussel, A., Gerbaud, G. and Courvalin, P. (1981) Mol. Gen. Genet. 182, 390-408.
13. Davis, R.W., Botstein, D. and Roth, J.R. (1980) in Advanced Bacterial Genetics. Cold Spring Harbor Laboratory, Ed., pp. 116-119, Cold Spring Harbor, New York.
14. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) in Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Laboratory, Ed., p. 170, Cold Spring Harbor, New York.
15. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
16. Staden, R. (1980) Nucl. Acids Res. 8, 3673-3694.
17. Gabarro-Arpa, J. and Michel, F. (1982) Biochimie 64, 99-112.
18. Ehrlich, R., Rodier, F., Gabarro-Arpa, J., Marin, M. Schmitt, B. and Reiss, C. (1981) CR Séances Acad. Sci. Paris [III], 1-4.
19. Wilbur, W.J. and Lipman, D.J. (1983) Proc. Natl. Acad. Sci. USA 80, 726-730.
20. Sellers, P.H. (1974) SIAM. J. Appl. Math. 26, 787-793.
21. Waterman, M.S., Smith, T.F. and Beyer, W.A. (1976) Adv. Math. 20, 367-387.
22. Smith, T.F., Waterman, M.S. and Fitch, W.M. (1981) J. Mol. Evol. 18, 38-46.
23. Tinocco, I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature New Biol. 246, 40-41.
24. Normore, W.M. (1976) Handbook of Biochemistry and Molecular Biology. Fasman, G.D., Ed. 3rd edn. Vol.II pp. 65-240, CRC Press, Cleveland, OH.
25. Rodier, F., Gabarro - Arpa, J., Ehrlich, R. and Reiss, C. (1980) C R Seances Acad. Sci., Paris [D], 199-202
26. Rodier, F., Gabarro-Arpa, J., Ehrlich, R. and Reiss, C. (1982) Nucl. Acids Res. 10, 391-402.
27. Hawley, D.K. and McClure, W.R. (1983) Nucl. Acids Res. 11, 2237-2255.
28. Ehrlich, R., Marin, M., Gabarro - Arpa, J., Rodier, F., Schmitt, B. and Reiss, C. (1981) CR Seances Acad. Sci. Paris [III], 5-8.
29. Grantham, R., Gautier, C., Gouy, M. Jacobzone, M. and Mercier, R. (1981) Nucl. Acids Res. 9, r43-r74.
30. Ikemura, T. (1981) J. Mol. Biol. 146, 1-21.
31. Nichols, B.P., Blumenberg, M. and Yanofsky, C. (1981) Nucl. Acids. Res. 9, 1743-1755.
32. Bibb, M.J., Findlay, P.R. and Johnson, M.W. (1984) Gene 30, 157-166.
33. Grantham, R., Gautier, C., Gouy, M., Mercier, R. and Pave, A. (1980) Nucl. Acids Res. 8, r49-r62.
34. Ikemura, T. and Ozeki, H. (1982) Cold Spring Harbor Symp. Quant. Biol. 49, 1087-1097.
35. Horinouchi, S., Byeon, W.H. and Weisblum, B. J. Bacteriol. 154, 1252-1262.
36. Carlier, C. and Courvalin, P. (1982) in Microbiology-1982, D. Schelessinger, Ed. pp162-166. American Society for Microbiology, Wasgington D.C.
37. Shaw, J.H. and Clewell D.B. (1985) J. Bacteriol. 164, 782-796
38. Vieira, J. and Messing, J. (1982) Gene 19, 259-268.
39. Novick, R.P., Clowes, R.C., Cohen, S.N., Curtiss III, R., Datta, N. and Falkow, S. (1976) Bacteriol, Rev. 40, 168- 189.